



The illusion of conscious experience

François Kammerer¹ 

Received: 27 July 2018 / Accepted: 19 December 2018
© Springer Nature B.V. 2019

Abstract

Illusionism about phenomenal consciousness is the thesis that phenomenal consciousness does not exist, even though it seems to exist. This thesis is widely judged to be uniquely counterintuitive: the idea that consciousness is an illusion strikes most people as absurd, and seems almost impossible to contemplate in earnest. Defenders of illusionism should be able to explain the apparent absurdity of their own thesis, within their own framework. However, this is no trivial task: arguably, none of the illusionist theories currently on the market is able to do this. I present a new theory of phenomenal introspection and argue that it might deal with the task at hand.

Keywords Consciousness · Illusionism · Introspection · Metaphysics · Physicalism

1 Introduction

Illusionism about phenomenal consciousness is the thesis that phenomenal consciousness does not exist, even though it seems to exist. This view has found some prominent defenders in contemporary philosophy, though it remains a minority view. One of its main advantages comes from the fact that it seems extremely difficult to locate phenomenal states in the physical world. By denying the reality of phenomenal consciousness, illusionists avoid the need to explain how physical processes can give rise to phenomenal states (the so-called “hard problem of consciousness”). They thus gain a way of defending physicalism about the human mind—an attractive metaphysical position threatened mainly by the fact that the phenomenal aspect of the mind seems precisely left out by the physicalist picture of the world.

According to illusionists, people usually believe—falsely—that they are phenomenally conscious because they are the victims of an illusion: the illusion of phenomenality. This illusion is thought to be an *introspective* illusion, and one way

✉ François Kammerer
kammerer.francois@gmail.com

¹ Institut Supérieur de Philosophie, FNRS, Université Catholique de Louvain, 14, place du Cardinal Mercier, 1348 Louvain-la-Neuve, Belgium

to describe it is to use the model of paradigmatic *perceptual* illusions, such as optical illusions. But a difficulty then arises for illusionists when we notice that the illusion of phenomenality seems quite peculiar. Indeed, the illusion of phenomenality is uniquely strong: it is extremely hard, not only to come to *believe* that phenomenality is illusory, but even to simply *contemplate* it in a clear and intuitive way. Merely representing to ourselves that phenomenal consciousness is an illusion creates a deep difficulty. No such difficulty attends perceptual illusions.

It is the burden of a good illusionist theory to explain why the illusion of phenomenality is so peculiar. I think that this is no trivial task. I intend to present a theory of phenomenal introspection which might be able to explain why the illusion of phenomenality is so unique. I call this theory the TCE theory. It states that phenomenal introspection, the specific process by which we come to form beliefs in a non-inferential way about our own phenomenal experiences, consists in the application of phenomenal concepts, which are Theoretically determined Concepts of Epistemologically special states (hence the acronym). According to TCE theory, there is an intimate connection between phenomenal introspection and our naïve epistemology, which is itself a part of our naïve theory of mind, in the sense that we systematically grasp phenomenal states as epistemologically special states. I argue that this may explain why the illusion of phenomenality is so peculiar and so hard to contemplate as such.

In §1, I present illusionism about conscious experience. In §2, I describe what I take to be one of the most—perhaps *the* most—important difficulties for illusionist views of consciousness, which I call the “illusion meta-problem”. I then describe the TCE theory (§3), and show why this theory might solve the illusion meta-problem (§4).

2 Illusionism about conscious experiences

Conscious experiences (“phenomenal states”, “phenomenal experiences”) are putative mental states such that there is “something it is like” to be in them. Seeing a red patch, tasting a chocolate cake, and feeling pain, are typical alleged examples of conscious experiences. These mental states are said to possess “phenomenal properties”, which characterize what it is like to be in these states. If a subject can enter such mental states, she is phenomenally conscious and possesses phenomenal consciousness.

Most philosophers admit that phenomenal consciousness is in *prima facie* tension with physicalism. The tension comes out through a variety of now classical considerations and arguments (Chalmers 1996; Jackson 1982; Kripke 1980; Levine 1983; Nagel 1974). Some philosophers (Chalmers 1996, 2010; Gertler 2008; Goff 2017; Jackson 1982; Nagel 2012) have taken at least some of these arguments to be successful, and stated that phenomenal consciousness is real and irreducible to anything physical—at least, if we don’t considerably reform our understanding of the physical—so that physicalism is false. However, the main disadvantage of such views is precisely that they force us to renounce physicalism, an otherwise very attractive metaphysical position.

Most philosophers have chosen to stick to physicalism regarding phenomenal consciousness. Some (Armstrong 1980; Dretske 1995; Lewis 1983a, b) have tried to show how consciousness can be reduced to some kind of physical process (broadly

understood as to include physically realized process) in a way that would make such reduction transparent—phenomenal truths being a priori deducible from physical truths. However, none of these attempts has reached general agreement within the philosophical community. Others have stated that, although phenomenal consciousness is entirely physical in nature, the reduction of consciousness to physical processes can never be transparent: phenomenal truths are not a priori deducible from physical truths. This last conception has been labelled “type-B” physicalism by Chalmers (2002) (while theories of Armstrong, Dretske and Lewis were put under the label “type-A” physicalism). Most (though not all)¹ defenders of type-B physicalism have also added that the reduction of phenomenal consciousness will not only remain opaque (in the sense that we will lack an a priori derivation of phenomenal truths from physical truths), but that consciousness will also persistently *seem distinct* from physical processes and irreducible to them (Aydede and Güzeldere 2005; Balog 2012; Kriegel 2009; Loar 1997; Papineau 2002; Sturgeon 1994). The most discussed attempt in that direction in recent years is the so-called “Phenomenal Concept Strategy” (for the term, see Stoljar 2005), which aims at explaining the apparent irreducibility of phenomenal consciousness by the special nature of phenomenal concepts. Many philosophers have adopted a view of this kind. However, there has been recently a growing concern that they cannot ultimately succeed: that it is impossible to maintain that physicalism is true of phenomenal consciousness and at the same time to account for our distinct epistemic situation regarding phenomenal consciousness in virtue of which phenomenal consciousness seems persistently distinct from the physical (Chalmers 2007; Demircioğlu 2013; Goff 2011; Levine 2001, 2007; Nida-Rümelin 2007). This point is still very much debated today (Diaz-León 2014; Elpidorou 2013, 2016; Schroer 2010; Shea 2014), but the idea that we really can “have it both ways” (Loar 1997, p. 598)—that is, account at the same time for the ontological reducibility of consciousness and for its persistent epistemic irreducibility—might now seem somewhat less plausible than it once did.²

The difficulties encountered by these forms of physicalism may incite physicalists to pay attention to a radical and counter-intuitive conception, according to which the human mind is entirely physical while phenomenal consciousness *simply does not exist*: there are no phenomenal properties instantiated in reality, and there is *nothing it is like* to be in any of our mental states. *Eliminative physicalism*, or *eliminativism* for short, has the advantage that it discharges our physicalist theories of the mind of the need to *explain* phenomenal consciousness (or to explain its merely apparent irreducibility) in physical terms. Indeed, eliminativists can recognize the radical irreducibility of phenomenal consciousness without creating difficulties for physicalism, given that for them phenomenal consciousness is nowhere to be found in reality.

Eliminativists can be distinguished by the way in which they propose to account for the fact that most of us falsely believe they are phenomenally conscious. Some eliminativists interpret this false belief as the result of a theoretical error—a kind of *doxastic mistake*. This is probably the case of some of the earlier proponents of

¹ See for example Block and Stalnaker (1999) and Yablo (1999).

² I will not argue here against the Phenomenal Concept Strategy. I have provided arguments against some of its most popular versions elsewhere (Kammerer 2018b, c).

eliminative materialism (Feyerabend 1963; Rorty 1965), and maybe also partly of the earlier Dennett (1988). We may call views of this kind “theoretical eliminativism”. Alternatively, some eliminativists insist that phenomenal consciousness is not merely a theoretical posit; it is something which persistently and robustly *seems* to exist. In this kind of view, the belief in phenomenal consciousness is not the result of a doxastic mistake; it rather stems from a kind of *introspective illusion*. We can call this view “illusionism”, which we may formulate as the conjunction of (i) the thesis that phenomenal consciousness does not exist (eliminativism) and (ii) the thesis that phenomenal consciousness nevertheless persistently *seems* to exist in a robust way (so that this “seeming” is unlikely to disappear on reflection or through the acquisition of new beliefs). I take it to be the most plausible version of eliminativism, and it is the view I am now going to focus on.³

Illusionists do not have to give a physical explanation of phenomenal consciousness, but merely a physical explanation of the fact that phenomenal consciousness *appears* to exist—that it seems to us that we are phenomenally conscious when in fact we are not. As Keith Frankish puts it: “Illusionism replaces the hard problem with the illusion problem—the problem of explaining how the illusion of phenomenality arises and why it is so powerful” (Frankish 2016, p. 37).

Three general approaches to this task can be found in the literature, appealing to (a) built-in, hard-wired features of our introspective mechanisms (Graziano 2013; Humphrey 2011; Pereboom 2011),⁴ to (b) a kind of mistaken inferential mechanism of projection (Rey 1995, 2007), or to (c) a mix of features of our introspective mechanisms and of philosophical (mostly Cartesian) prejudices (Dennett 1988, 1991, 2017). For my part, I believe (a)-style theories are best positioned to account for the *robustness* of the illusion of phenomenality: they can explain more easily the fact that this illusion does not disappear or diminish on reflection. This robustness of the illusion of phenomenality will be a central concern in this paper.

One cannot deny that illusionism (as any form of eliminativism) is a deeply counter-intuitive view. However, it does not have *all* of the counter-intuitive consequences it is sometimes thought to have. For example, illusionists are *not* committed to the view that our introspective states (such as the phenomenal judgment “I am in pain”) do not reliably track any *real* and important psychological property. They simply deny that such properties are *phenomenal*, and that there is something it is like to instantiate them. Frankish suggests calling such properties “quasi-phenomenal properties” (Frankish

³ The use of the term “illusionism” has been recently popularized by Frankish (2016), even though he does not distinguish as I do between eliminativism and illusionism. Frankish also makes other distinctions, for example between weak and strong illusionism. “Weak illusionists”, contrary to strong illusionists, believe that phenomenal consciousness exists. However, they also state that consciousness seems to have some properties that it does not really have. I take it that weak illusionism constitutes an unstable position, and that strong illusionism is the best version of illusionism (Chalmers 2018, pp. 49–52; Frankish 2012). Here what I call “illusionism” is closer to what Frankish calls “strong illusionism”.

⁴ It is to be noted that neither Humphrey nor Graziano accept the term “illusionism” to characterize their theories. Humphrey once accepted the characterization but now calls himself a “phenomenal surrealist”, as he feels this term captures best the importance we give to our representation of phenomenal properties (Humphrey 2016, 2017). Graziano (2016) rejects the term “illusionism” because he wants to limit the use of the vocabulary of “illusion” to rare and abnormal dysfunctions of a detecting mechanism. However, both are illusionists in the sense I defined.

2016, p. 15)—purely physico-functional and *non-phenomenal* properties which are reliably tracked (but mischaracterized as phenomenal) by our introspective mechanisms. For the same reason (Frankish 2016, p. 21), illusionists are *not* committed to the view that a mature psychological science will not mention any form of consciousness beyond, for example, access-consciousness. After all, quasi-phenomenal consciousness may very well happen to have interesting distinctive features from the point of view of a psychologist.

To sum up: physicalism is an attractive position concerning the human mind. Because it seems threatened by the existence of phenomenal consciousness, and because standard physicalist responses to this threat encounter various deep difficulties, some philosophers have been led to see eliminativism, and particularly illusionism—arguably, its most plausible version—as an interesting theoretical alternative to more standard versions of physicalism. Illusionism, of course, like any position, can be subjected to numerous objections, which find in turn some answers (for an overview, see Frankish 2016, pp. 29–37). However, I think that all current illusionist theories of consciousness face one major problem, which has often been under-estimated (if not simply ignored) by proponents of illusionism. I call it the “the illusion meta-problem”.

3 The illusion meta-problem

Of the many challenges facing illusionism, one stands out. I call it the “illusion meta-problem” (Kammerer 2018a), and I think that it constitutes the hardest aspect of the illusion problem.⁵ While the *illusion problem* is the general problem of explaining how the illusion of phenomenality arises, the *illusion meta-problem* concerns the explanation of its *peculiar strength*. More specifically, it is the problem of explaining not only why phenomenal consciousness seems to exist even though it does not (why we have an illusion of phenomenality), but also why it seems so strongly to exist (why the illusion of phenomenality is so peculiarly strong); notably, and that’s the crucial point, why it is so hard for us, not only to *believe* that phenomenality is an illusion, but even to seriously *entertain* the possibility that it is.

Let us elaborate on this point. If illusionists are right, then phenomenal consciousness is a kind of introspective illusion, and it seems that we should be able to think about it more or less on the model of other illusions, such as perceptual illusions. But there is a clear contrast between the alleged illusion of phenomenality and perceptual illusions. Once described, this contrast makes the illusion meta-problem salient.

Take the Müller-Lyer illusion. We can accept, without too much difficulty, after some thinking and some checking, that it is really an illusion: that our visual representation of the relative length of the lines is incorrect—even though the illusion does not *disappear* once we accept this. And the same is true of other paradigmatic perceptual illusions: after some thinking and some checking, we normally have no difficulty accepting their illusory character. However, most people find it almost impossible

⁵ David Chalmers calls the illusion meta-problem the “resistance problem” (Chalmers 2018, p. 27), in order to avoid confusion with his own “consciousness meta-problem”.

to do the same thing with phenomenal consciousness. There may be *arguments* in favor of illusionism; all the same, the idea that our introspective grasp of phenomenal states is illusory, so that there is *nothing it is like* to be in such states, simply seems insane and almost impossible to believe. Philosophers have called illusionism “crazy” (Frances 2008, p. 241; Strawson 1994, p. 101), “utterly implausible” (Balog 2016, p. 42), “impossible” and “absurd” (Nida-Rümelin 2016, p. 163, 170), “obviously false”, “self-defeating” and “incoherent” (Goff 2016, pp. 84–85), and other things in the vicinity (Chalmers 1996, pp. 188–189; Searle 1997).

One might suggest that illusionism seems so implausible simply because our introspective disposition to believe that we are phenomenally conscious is *stronger* than our perceptual dispositions to believe that such and such perceptual illusory situations are real. For example, it could be that my introspective disposition to believe that I have a phenomenal experience of red right now is stronger than my visual disposition to believe that the two lines in the Müller-Lyer illusion have different lengths. Likewise, it could be that our disposition to believe that consciousness is an illusion is simply weaker than in the case of perceptual illusions, given that we cannot directly *check* its illusory character—we simply appeal to indirect theoretical considerations.

However, the contrast between the illusion of phenomenality and other illusions goes deeper than that. For example, it is undeniable that I have an extremely strong perceptual disposition to believe that I have two hands right now. I am quite convinced that this perception of my two hands is *not* illusory. It would take a considerable amount of converging evidence (which I currently *do not have*) in order for me to stop believing, on the basis of my perception, that I have two hands right now. So: I have a very strong disposition to believe that I have two hands, and no disposition to believe that I do not, which is why it would be very hard for me to come to *believe* that I do not have two hands. However, and that’s the crucial point: I have no problem—even the slightest difficulty—*entertaining* the hypothesis that I do not have two hands, and that, for example, I am simply the victim of a hallucination. I have no problem *representing* to myself such a situation—in fact, I find it quite easy. This situation seems perfectly possible and coherent, and I have no difficulty forming a positive conception of it—even though I do not believe this situation to be the case and would certainly not believe this easily.

But the case of consciousness is quite different: we have trouble simply *entertaining* the situation described by illusionism about consciousness. When we try to consider the hypothesis that, for example, “it introspectively seems to me that I have an experience of red right now, but in reality I do not have an experience of red”, we find ourselves deeply puzzled: we have deep difficulties even conceiving clearly of this situation and forming a positive conception of it. It is not only that, in such a case, we have a very strong disposition to believe that we have this experience of red and that things are as they appear (counterbalanced by no disposition to believe the contrary); it is rather that we just cannot clearly represent to ourselves a situation in which it introspectively seems to us that we have an experience of red while we do not have this experience.⁶

⁶ I am in no way implying that we have a similarly deep trouble representing to ourselves a situation in which we commit a *phenomenal error*—i.e. a situation in which we make an *incorrect judgment* about our experience. What is really conceptually problematic for us is not phenomenal error, but *phenomenal*

What I call the illusion meta-problem is the problem of explaining this unique difficulty we encounter when we try to think of the introspective appearance of phenomenality as an illusion. My opinion is that none of the illusionist theories currently on the market is able to solve this problem.⁷

The illusion meta-problem is related to the notion that there is no appearance/reality distinction when it comes to consciousness (Kripke 1980; Searle 1997). This notion, indeed, could explain why situations in which it *appears* to us that we are conscious, when in fact we are not, seem simply *impossible*. The problem for illusionists is that they must ultimately hold that there *is* an appearance/reality distinction for phenomenal consciousness, given that they hold that phenomenal consciousness appears to exist without really existing. Therefore, one aspect of their challenge is to give an account of why it is so hard to draw a distinction between appearance and reality about phenomenal states, even though such a distinction holds.

One thing deserves to be noted. The main reason illusionism about consciousness tends to be rejected is that it seems crazy; but it is in turn quite plausible that the main reason it seems crazy is precisely that we encounter deep difficulties when we try to represent to ourselves that phenomenal consciousness does not exist even though it seems to exist. This, by the way, may explain why illusionist theories of consciousness so far have failed to recognize the importance of the illusion meta-problem. Indeed, it might be that, by recognizing the deep difficulty we encounter when we try to conceive of the illusory nature of consciousness, illusionists were afraid to grant too much to their phenomenal-realist opponents. After all, they were arguing *for* illusionism, and it is rarely a smart move in an argument to start by admitting that your own position persistently seems incoherent. This might have led proponents of illusionism to (more or less consciously) sweep what I have called the illusion meta-problem under the carpet. Whether or not this psychological hypothesis is true, I think that it is time for illusionists to face up to this problem. This could also reinforce their dialectical position: after all, most philosophers keep thinking that illusionism really is crazy. If illusionists want to make progress on the reproach of craziness, they have to take this fact seriously and provide an *explanation* of why their own position persistently seems crazy, even though it is true. There is no guarantee that providing such an explanation could completely undermine the reproach of craziness (Chalmers 1996, pp. 188–189), but illusionists would at least get some dialectical leverage in the debate against phenomenal realists. So, the fact that illusionism about phenomenal experiences is so hard for us to represent as such is itself a salient fact

Footnote 6 continued

illusion—i.e. a situation in which everything introspectively appears to us as if we had an experience, but we do not have this experience. There are cases that can be shown to be rather intuitive cases of phenomenal errors, such as the fraternity case discussed by Pereboom (2011, pp. 22–23) and Hill (1991, pp. 128–129): a student about to be initiated in a fraternity is blindfolded and told that a razor is about to cut his throat. He feels something on his throat and judges that he is in pain; but then realizes that what he feels is simply a sensation of cold, and that an icicle has been put on his throat instead of a razor. When we think of such a situation, we can intuitively think of it as a case of *phenomenal error*. What is really difficult, however, is to think of this situation (or of other situations) as a case of *phenomenal illusion*: as a situation in which *everything appears to the subject as if he had an experience of pain*, while he does not have this experience, even for a short moment. I argued against interpretations of similar cases example as intuitive cases of phenomenal illusions in Kammerer (2018a, pp. 58–61).

⁷ I argued extensively for this elsewhere (Kammerer 2018a).

about the illusion of phenomenality that illusionists aim at explaining; therefore, it is a legitimate part of the *explanandum* of their theories. But illusionists should also be able to explain this fact for dialectical reasons: they need to account for their own apparent craziness, precisely to undermine their opponents' reproach.

4 Phenomenal introspective representations as theoretically determined concepts of epistemologically special states

I now intend to present an illusionist theory of phenomenal introspection aimed at solving the illusion meta-problem: TCE theory.⁸ In this section, I will present TCE theory in a dogmatic way; the next section will be devoted to arguing that it solves the illusion meta-problem. I don't intend here to argue directly for the truth of this theory, nor for the truth of illusionism in general, but simply for the following conditional: if TCE theory correctly describes how phenomenal introspection works, then we should expect illusionism regarding consciousness to be uniquely hard to represent to ourselves as such, even though it is true. In other words, TCE theory, if true, *solves the illusion meta-problem*. If this theory really succeeds at solving the illusion meta-problem, it is of crucial interest to proponents of illusionism. However, it can also be of interest to phenomenal realists who are curious to find out what the best version of eliminativism might be.

The TCE theory is a theory about how the mind works. As such, its formulation belongs to the domain of mere speculative philosophical psychology. In order to argue for the *truth* of such theory, one would have to gather empirical evidence supporting it. I am not going to provide such empirical evidence here (even if I will point at the coherence of the TCE theory with some empirical research programs in the cognitive science). My goal is merely to provide a "how possibly explanation" of the unique difficulty we face when we consider illusionism about consciousness. I here share the point of view of Chalmers (2018, p. 10): in the nascent multidisciplinary research program aiming at the explanation of our intuitions about consciousness, the role of philosophy should be to assess potential mechanisms underlying phenomenal reports. This paper intends to describe and to assess such a mechanism, described at a high level of abstraction.

A *caveat*: TCE theory, as developed here, is intended to be a theory of introspection of *perceptual* phenomenal states. It is silent on introspection of non-perceptual phenomenal states, such as bodily phenomenal states (bodily sensations), algedonic phenomenal states (phenomenal pain, phenomenal pleasure, etc.), or such as hypothetical cognitive phenomenal states [whether or not we introspect such states being itself a hotly debated question (cf. Bayne and Montague 2011)]. I focus my account on the introspection of perceptual phenomenal states, and I leave open the question of knowing whether or not this account could be extended, *mutatis mutandis*, to the

⁸ Again, by "phenomenal introspection" I mean the specific process by which we come to form beliefs in a non-inferential way about our own phenomenal experiences.

introspection of bodily phenomenal states, algedonic phenomenal states, or to the introspection of hypothetical cognitive phenomenal states.⁹

According to TCE theory, phenomenal introspection consists in the application of phenomenal concepts, which are Theoretically determined Concepts of Epistemologically special states (TCE). These concepts are governed by our naïve (“folk”) and modular *theory of mind*, which includes a naïve *theory of knowledge*, that is, a naïve epistemology. In this view, there is a tight link between the introspection of phenomenal states and our naïve epistemology; this tight link, I will suggest, is what accounts for the uniqueness of the illusion of phenomenality.

TCE theory falls within the “theory–theory of self-awareness” approach. The basic idea of this approach is that the introspective access one has to one’s mental states depends on the same capacities and mechanisms that are used to attribute mental states to others—and that these capacities include a body of information about mental states—a theory of mind.¹⁰ As I see it, introspection is a theoretically informed activity, governed by our naïve theory of mind. This theory has to be understood as a set of capacities that allow human beings to describe, explain and predict the behavior of creatures (including themselves) by reference to mental entities.¹¹

According to TCE theory, this theory of mind is not explicitly constructed and/or applied at the personal level; it functions at a sub-personal level. This point is important, as it implies that, even if this theory of mind constrains and determines our judgments and reasoning about the mind, we should not expect the “principles” of such theory to be easy to articulate at a personal level. Plausibly, this theory of mind is implemented in a *modular* way. The notion of a cognitive module was first introduced by Fodor (1983). Here I have a rather weak understanding of modularity: by “module”, I mean a cognitive subsystem which is informationally encapsulated to a relevant degree, is partially inaccessible to the “central system”, is domain-specific, innate, and has a quick and partly automatic functioning (Scholl and Leslie 1999, pp. 133–134).¹²

⁹ For some elements in favor of such an extension, see Kammerer (2016a, b, pp. 241–242).

¹⁰ One can find a definition and a critical overview of the “theory-theory of self-awareness” approach in (Nichols and Stich 2003, pp. 164–169). This kind of approach, broadly understood, is embraced by numerous researchers (Aydede and Güzeldere 2005; Carruthers 1996, 2005, 2011; Frith 2002; Frith and Happé 1999; Gazzaniga 1970; Gopnik 1993; Gopnik and Meltzoff 1994; Graziano 2013; Nisbett and Wilson 1977; Reuter 2013, 2014).

¹¹ Many philosophers of science (following Hanson 1958) have talked about the “theory-ladenness” of scientific observation—meaning that the content of scientific observations is at least partially determined by the theories embraced by the scientists doing the observations. Churchland (1979), drawing on such views, has claimed that first-person ascriptions of mental states were theory-laden in a similar sense. When I say that introspection is a theoretically informed activity, my view bears some similarity to his, except that I crucially claim (more on this shortly) that the theory of mind which determines introspection is not a theory we can choose to embrace or not: it functions at the sub-personal level and takes the form of an innate module.

¹² A subsystem is informationally encapsulated and partially inaccessible to the central system when there are important restrictions in the information that can flow from the rest of the system to the subsystem (encapsulation) or from the subsystem to the central system (partial central inaccessibility). A subsystem is domain-specific when the class of objects and properties it processes information about is relatively narrow.

According to TCE theory, introspection consists in the application, governed by our theory of mind module, of phenomenal concepts.¹³ These phenomenal concepts are themselves “hybrid” concepts (roughly of the type described in Frankish 2016, p. 36–37). They have two components: one component is a recognitional concept which refers to detectable properties of *external objects* (or of the body), the other component is a general theoretical operator common to all phenomenal concepts. For example, let us take the phenomenal concept of experiences of red (call it ⟨experience of red⟩¹⁴). This concept has a sensory, recognitional concept ⟨red⟩ as one of its components. This recognitional component refers to the sensible property *red*, which is a property possessed by some surfaces; and this recognitional concept ⟨red⟩ thus refers thanks to its link to some recognitional, sensory capacity sighted people possess. This sensory concept ⟨red⟩ (like other basic sensory concepts) presents the property of redness in a direct, unanalyzable way—as a primitive property of surfaces. The other component of the concept ⟨experience of red⟩, which is common to all phenomenal concepts, is a general theoretical operator ⟨experience⟩, the content of which is determined by our theory of mind module (more on this shortly). So, ⟨experience of red⟩ is our phenomenal concept of experiences of red, and it is formed by the sensory concept ⟨red⟩ and the conceptual operator ⟨experience⟩. And all phenomenal concepts are thus formed by two such components.

The content of our naïve concept of experience—the invariant component of all our phenomenal concepts—is a *theoretical* content: it is determined by the overall content of our theory of mind module. Naturally, we cannot here reproduce this entire theory. But the core of the theory can be conveniently described through the following four core principles.¹⁵

- (1) Principle of Appearance: certain states of affairs can *appear* to subjects. An appearance can be either veridical or correct (when the appearing state of affairs is the case), or nonveridical or incorrect (when it is not)
- (2) Principle of Receptive Affection: states of affairs appear to subjects in virtue of internal receptive affections of the subjects, which constitute *experiences*
- (3) Principle of Resemblance: what a given receptive affection makes appear to a subject is determined by what this receptive affection maximally resembles to
- (4) Principle of Individuation: an experience of X is a receptive affection of a subject which makes X appear to the subject

What the *Principle of Appearance* (1) states is that at least some (and possibly all) subjects who enter mental states can be in states of *appearance*, in which a certain state of affairs *appear* to them—correctly (veridically) if the state of affairs obtains, incorrectly (nonveridically) otherwise. The principle does not imply that *all* states of affairs can thus appear—the question of knowing which states of affairs are considered capable of such appearances by our theory of mind is beyond the scope of this paper.

¹³ So, phenomenal introspective representations are conceptual. I think that one could hold a quite similar view, in which phenomenal introspective representations are seen as pre-conceptual (even though a few complications would arise).

¹⁴ Following a widespread convention, I write “⟨x⟩” to refer to the concept of x.

¹⁵ Our theory of mind does not consist in a set of stated principles, but in a set of representational and inferential capacities. However, it is quite convenient to present it under the form of principles.

The *Principle of Receptive Affection* (2) expresses what our theory of mind states about the internal basis, in subjects, of these appearances. The idea is that subjects can enter *internal* states (which do not depend constitutively on their environment); some of these states can be states of *receptive affection*, which means that they do not constitutively depend, for their existence, on any bodily or mental “action” of the subjects (by mental actions here I mean judging, deciding, and so on—mental processes which are seen by our naïve theory of mind as being under the direct control of the subject). These receptive affections, according to our theory of mind, are what ground the appearance of states of affairs to the concerned subject: they make these states of affairs appear to the subjects. In ordinary and philosophical talk we call such internal receptive affections “experiences”, and the concept (experience) is the concept of a receptive affection in this sense.

The *Principle of Resemblance* (3) states that the way in which subjects are receptively affected determines what appears to them. The fact that a subject is receptively affected allows a certain state of affairs to appear to the subject in virtue of the fact that the way in which the subject is receptively affected *resembles in a certain way, in the mental domain, this state of affairs*. More precisely: *a state of affairs X appears to a subject S if and only if S is receptively affected in a way that maximally resembles X*. Here is what I mean by “maximally resembles”: a receptive affection A of a subject S maximally resembles a state of affairs X if and only if (i) A resembles X, i.e., A “reproduces” X (or “corresponds” to X) in the mental domain; (ii) A is the receptive affection that resembles X the most (or one of the receptive affections that equally resemble X the most), amongst all the ways in which S can be receptively affected.

The *Principle of Individuation* (4) states that we individuate experiences by what they make appear to subjects: an “experience of X” is a receptive affection of a subject which makes X appear to the subject. However, only the state of affairs that a given experience makes appear *and which is not identical with the tokening of the experience itself* can thus individuate the experience. This caveat is important because, as will be made clear, the content of our theory of mind has the consequence that experiences make themselves appear to subjects. Aside from that, the Principle of Individuation has a notable consequence: because experiences are individuated by the appearances they ground, they too can be said to be correct (veridical) or incorrect (nonveridical), depending on the corresponding status of the appearances they ground.

A few things should be stressed about (1–4). First and foremost, I do not intend to use these statements to describe the nature of phenomenal experiences themselves, but merely *our naïve theory of mind module’s take* on phenomenal experiences. Crucially, TCE theory is not at all committed to the view that there *really is* a relation of resemblance between our experiences and the states of affairs that experiences make appear (Principle of Resemblance). All it is committed to is the view that our theory of mind module *does posit implicitly* such a relation of resemblance, and *does define implicitly* what experiences are through this relation of resemblance.

Secondly, it should be clear that the theoretical content of phenomenal concepts expressed by these statements is not itself directly accessible at a personal level to the subjects manipulating these concepts. This content, in fact, need not be explicitly represented as such anywhere; it may well be grounded in the inferential role played by

phenomenal concepts—notably, in the inferential relations of phenomenal concepts with other concepts of our theory of mind module, such as the concept of appearance.

The view I described so far is but a theory of phenomenal concepts; it says nothing about phenomenal consciousness in itself. As such, it is committed neither to realism nor to eliminativism about phenomenality. However, as I intend to use this view to solve the illusion meta-problem, I suggest combining it with eliminativism. Combining this account of phenomenal concepts with eliminativism means that our phenomenal concepts are not satisfied, because nothing satisfies the theoretical characterization of experiences contained in the conceptual operator ⟨experience⟩. Nothing in reality is such that it is a receptive affection which makes things appear to a subject in virtue of its maximal resemblance to the appearing thing.¹⁶ However, it still appears to us (in a sense of appearance that will be explained below) that we are in states of receptive affections so characterized, which makes TCE theory an illusionist theory of consciousness.¹⁷

As noted earlier, I do not plan to argue for the truth of TCE theory in this paper, but simply for the idea that this theory has the resources to solve the illusion meta-problem. However, let me address briefly the potential complaint that the theory is entirely ad hoc—a hypothesis without independent support, tailored to defend a doctrine (illusionism) against some recalcitrant facts (the illusion meta-problem). The force of such criticism cannot be properly assessed without carefully considering the arguments that could be put forth in favor of TCE, but note that TCE theory does not make any extravagant assumptions—assumptions that cannot be judged plausible for independent reasons. Thus, (1) TCE theory relies on the idea that introspection is a theoretically-determined process. Even if such an idea is rejected by many, it is also embraced for independent reasons by the numerous researchers working in the perspective of the “theory–theory of self-awareness”. (2) TCE theory also relies on the thesis that our naïve theory of mind conceives of the way in which experiences ground what appear to the subject in a manner that appeals to the *resemblance* of these expe-

¹⁶ In other words, this means that there is nothing in reality that satisfies together the four principles of our naïve theory of mind presented earlier. As an anonymous reviewer points out, principles (1), (2) and (4) could plausibly be satisfied by something purely physical. Arguably, a satisfying physicalist theory of the mind would vindicate the idea that states of affairs can appear to subjects in virtue of the internal states of these subjects (and, trivially, that these internal states can be individuated by the states of affairs they make appear). For that reason, the kind of eliminativism I suggest here concerns exclusively entities that would not only satisfy principles (1), (2) and (4), but also, crucially, principle (3)—the Principle of Resemblance. What I deny is that there are internal states of subjects which make states of affairs appear to subjects if and only if they maximally resemble these states of affairs.

¹⁷ The way I understand it, terms such as “phenomenal consciousness” express a concept which corresponds to the conceptual operator (experience). For this reason, if this concept is not satisfied (as I claim), phenomenal consciousness does not exist. That makes my position a variety of *strong illusionism*. If weak illusionists were to use the TCE, on the other hand, they would use the term “phenomenal consciousness” differently: they would state that phenomenal consciousness *does exist* though it does not have the properties it seems to have—notably, it does not satisfy our conceptual operator (experience), though it seems to satisfy it. A position of this kind would, as a weak illusionist one, fall under the category “type-B physicalism” (according to which consciousness is reducible to some physical process, even though this reduction will always remain opaque, and maybe also deeply counter-intuitive), while strong illusionism is more naturally associated with type-A physicalism. As such, a position of this type would encounter the problems which are typical of type-B physicalism in general, and of weak illusionism in particular (Chalmers 2018, pp. 49–52; Frankish 2012).

riences with the appearing states of affairs. This idea may be unacceptable to many philosophers, but it must be kept in mind that TCE theory does not say that there *is* such a relation of resemblance between an experience and what this experience makes appear; merely that our naïve theory of mind says so. And this last idea is itself, I think, extremely plausible in its own right. Indeed, we find numerous versions of the idea that there is a resemblance between our experiences of things and things themselves (in virtue of which our experiences present these things) throughout the history of philosophy,¹⁸ which in turn supports the thesis that this idea is deeply intuitive and part of our naïve understanding of experiences. (3) Finally, although TCE theory has the consequence that our naïve theory of the mind is deeply inaccurate, the notion that one of our naïve theories is deeply inaccurate does not seem problematic in itself. Our naïve physics, for example, contains plenty of fundamental errors concerning the kind of physical entities there are in the world. It includes the false idea that there are basic forces such as “sucking” (used by vacuum-cleaners), or that the movement of physical objects is caused by a kind of internal “impetus” (Hayes 1978; McCloskey 1983). Similar things could be said about our naïve zoology, our naïve sociology, and so on.

5 Solving the illusion meta-problem

How does TCE theory solve the illusion meta-problem? That is, how does it account for the fact that illusionism regarding conscious experience is so hard, not only to accept, but even to simply represent to ourselves? My claim will be that, if introspection really works as TCE theory states, this has the consequence that, through introspection, we grasp experiences as entities that cannot introspectively appear in a nonveridical way; and this, in turn, solves the illusion meta-problem. I will now argue for this idea, and try to make it clear.

Start with an example. Suppose we judge that Julie has an experience of a red rose in front of her. Given TCE theory’s view of the content of our phenomenal concepts, this means that we (implicitly) judge that she is receptively affected in a way which maximally resembles the presence of a red rose in front of her. For it is the receptive affection that resembles the most the presence of a red rose in front her, amongst all the ways in which she can be receptively affected.¹⁹ And we implicitly judge that it is in virtue

¹⁸ We can for example think of Aristotle’s theory of perception developed in the *De Anima*, which influenced most of the western medieval conceptions of perception, or of Epicurus theory of perception, which posits the existence of *simulacra* of things as the basis of perceptual states.

¹⁹ It could be the case that we judge Julie to be receptively affected in a way that *resembles* the presence of a red rose in front of her, but that is not the way that *resembles the most* this state of affairs. For example, let’s consider the receptive affections that we would call an “experience of an orange rose” or an “experience of a pink rose”. Our theory of mind would characterize them as resembling the presence of a red rose, but not as being the affections that most resemble the presence of a red rose (this would be an experience of a red rose). On this matter, two things should be remarked. First, when it comes to states of affairs that are not *concretely fully determined*, but are partially abstract (for example, let’s not consider *the presence of red rose with such shape and such hue and such distance* but rather *the presence of red*), there will be *many different affections* that will be judged to equally maximally resemble this state of affairs, so that it will be impossible to determine a *unique* kind of affection that is “truly” (and uniquely) an experience of red. Therefore, an experience of a red rose, an experience of a red car, an experience of a red circle, etc., can all be called “experiences of red”. And it is only for concretely fully determined states of affairs that we can

of this resemblance that this receptive affection makes the red rose appear to Julie. If, when the mind of Julie is thus receptively affected, there is indeed a red rose in front of her, then we judge her experience to be veridical—otherwise, to be nonveridical. In the latter case, we judge that Julie has an experience of a red rose—but an illusory one.²⁰

Here we understand how our theory of mind module allows us to intuitively represent, via phenomenal concepts, how Julie can have a nonveridical appearance of the presence of a red rose in front of her—an illusion of a red rose. Indeed, our theory of mind, thanks to phenomenal concepts, has the representational resources to describe this situation of nonveridical appearance of a red rose, and can describe what state of the subject grounds this appearance. In fact, one of the main features of our naïve concepts of *appearance* and *experience* is precisely that they allow us to describe, interpret and theorize such situations. Our concept of experience allows us to intuitively explain how a subject can enjoy an appearance of a thing whether or not the thing is present: because the receptive of affection of the subject, which makes the thing appear to the subject, is held constant.

But now consider the particular case of appearances of *experiences* themselves, that is, of *introspective* rather than *perceptual* appearances. According to our theory of mind module, can the fact that one has a certain experience—for example, an experience of a red rose—appear to the relevant subject? Let us take Julie, for example. In order for her to be in a situation such that it will appear to her that she has an experience of a red rose, she will have to be receptively affected in a way that maximally resembles the state of affairs in which she has an experience of a red rose. But here, a peculiar difficulty arises. An experience of a red rose *is after all itself a certain state of receptive affection*. But this notably and crucially implies that, amongst all the ways in which Julie can be receptively affected, the way which *resembles the most* an experience of a red rose *will itself be an experience of a red rose*. Indeed, for any given thing, it is a trivial truth that nothing resembles that thing as much as the thing itself.

Now, this has two major consequences. The first consequence, and the most important for our purposes, is as follows. If it appears to Julie that she has an experience of a red rose, this means that she is receptively affected in a way that maximally resembles an experience of a red rose. Given that no receptive affection resembles as much a given receptive affection of a type A than a receptive affection of the exact same type

Footnote 19 continued

expect that one will be able to determine a single kind of receptive affection that constitutes an experience making this state of affairs appear. Second, we sometimes group together experiences simply because we take it that they are caused by a single kind of object. For example, we can talk of “experiences of roses”. But of course, one can have a visual experience of a rose, an olfactory experience of a rose, a tactile experience of a rose, etc. And if all these experiences, according to our theory of mind, *maximally resemble* some state of affairs when taken individually (states of affairs respectively described with sensory visual concepts, sensory olfactory concepts, sensory tactile concepts), it is unlikely that we judge that there is a unique state of affairs that *all of them happen to maximally resemble*. This means that, according to our theory of mind, there is no such thing, from an experiential point of view, as an “experience of a rose” *simpliciter*: such an expression does not correspond to any real experiential kind. However, this does not contradict the fact that we *talk* easily enough about experiences of roses. It simply means that we linguistically construct an expression that refers to a disjunction of what our theory of mind may consider as “real” experiential kinds (visual experiences of roses, olfactory experiences of roses, etc).

²⁰ In what follows, I do not draw any distinction between illusions and hallucinations, and I just take “illusion” to mean “nonveridical appearance”.

A itself, this means that she has an experience of a red rose. So, according to our theory of mind, if it *appears* to Julie that she has an experience of a red rose, then she *has* an experience of a red rose. This point can obviously be extended to all experiences. Therefore, according to our naïve theory of mind, which—recall—determines the content of our phenomenal concepts, if it appears to a subject that she has an experience, then necessarily she has this experience. Accordingly, introspective appearances of our experiences are necessarily always correct. In other words, TCE theory predicts that our naïve theory of mind implies that introspection is in a certain sense infallible.

The second consequence is as follows. If Julie has an experience of a red rose, then she is receptively affected in a certain way. The way in which she is affected happens, in fact, to *resemble* an experience of a red rose, and even to *maximally resemble* such an experience (again, it is another trivial truth that, if any two receptive affections are identical, then those two receptive affections maximally resemble each other). So, according to our theory of mind module, if Julie has an experience of a red rose, then it appears to her that she has an experience of a red rose. This point, again, can be extended to all experiences. Therefore, according to our naïve theory of mind, which—again—determines the content of our phenomenal concepts, if a subject has an experience, then it appears to her that she has this experience. In other words, TCE theory predicts that our naïve theory of mind implies that *experiences are self-intimating*.

In this way, TCE theory predicts that our naïve theory of mind is committed to the two familiar tenets of Cartesian conceptions of self-knowledge: infallibility and self-intimation.

If we examine the first consequence just described (regarding infallibility), it should be clear that the content of phenomenal concepts, as determined by our theory of mind module, makes it impossible for us to think that it appears to us that we have an experience even though we don't have the experience—at least as long as the concept of appearance we then use is our naïve concept, the concept provided by our theory of mind module (more on this later). Any time we think about an appearance of an experience to a subject, we think about something that is a receptive affection which maximally resembles this experience, so that it has to *be* an experience of the same kind. For this reason, there cannot be a *nonveridical* appearance of an experience—this simply is a contradiction, given the content of our concepts. Note that this problem only arises in the case of experiences: other things can *appear* to a subject without really *being* there. But this explains why we encounter a unique difficulty when we try to represent to ourselves the illusory character of phenomenal experience—it explains why there is something uniquely *incoherent-seeming* in illusionism about consciousness.

However, we should realize that this incoherence only arises to the extent that we use our naïve concept of appearance—the concept of appearance which is provided by our naïve theory of mind, and which is conceptually linked to our phenomenal concepts. But other concepts of appearance can be constructed. For example, one could construct a *functional* and scientific concept of appearance, which would define an appearance as a momentary and non-cognitively penetrable disposition to believe something (“belief” being defined in a purely functional way too). If we use such a scientific concept of appearance, then there is no contradiction, no incoherence, in the idea that phenomenal experiences appear to exist, but do not really exist. Of course, even if we manage to

prevent ourselves from using our intuitive concept of appearance (which is no trivial task) and simply use our functional and scientific concept, illusionism will not cease to be counter-intuitive: after all, we are still provided with *appearances* (functionally conceived) of experiences by introspection, which dispose us to believe that we really have phenomenal experiences. But illusionism ceases to be *uniquely* problematic and to seem incoherent: the thesis that consciousness is illusory, so understood, acquires a status similar, for example, to the thesis that colors, understood as primitive and uncomposed properties of surfaces, do not exist but simply seem (perceptually) to exist.²¹

So: TCE theory explains why the situation described by illusionism is, in a way, *unthinkable*: it leads to a contradiction when we try to think it while using our intuitive and naïve concept of appearance. This in turn explains why we cannot intuitively make sense of illusionism; it explains why we struggle so much to represent to ourselves, in an intuitive fashion, that consciousness seems to exist but does not exist.²² It explains why such a problem only arises for illusionism about consciousness (and not about other entities). Moreover, it also explains why illusionism in itself is neither contradictory nor incoherent at all: as long as we use a functional (but non-intuitive) concept of appearance, illusionism makes perfectly good sense. This is how TCE theory solves the illusion meta-problem.^{23,24}

²¹ This kind of view regarding colors has been defended throughout the history of philosophy, at least since Galileo and Descartes and maybe since the works of ancient Atomists. Many contemporary philosophers have defended similar views (Chalmers 2006; Hardin 1988; Maund 2006). The analogy between illusionism regarding consciousness and illusionism regarding primitive colors is used, in defense of illusionism regarding consciousness, by Pereboom (2011). One problem of his view, in my opinion (Kammerer 2018a), is that it precisely fails to explain why illusionism regarding consciousness poses some peculiar extra difficulty when compared with illusionism regarding primitive colors.

²² Which explains naturally why saying that phenomenality is an illusion is “the sort of thing that can only be done by a philosopher—or by someone else tying themselves in intellectual knots” (Chalmers 1996, p. 188). But of course, in my view, this does not count as a criticism of illusionism.

²³ TCE theory bears some similarity in spirit to a view I defended in previous work (Kammerer 2016c). If I set aside the fact that TCE is developed in more details, the two main differences are the following. (1) In my previous view, the intuitive impossibility of nonveridical appearances of experiences stems from the fact that our naïve theory of the mind posits that, in order for something to be a nonveridical appearance of X, it has to be a state entirely similar, from an experiential point of view, to a true appearance of X. However, this can easily come across as an arbitrary supposition. TCE theory shows how the intuitive impossibility of nonveridical appearances of experiences can itself be seen as consequence of a more general principle, which our theory of mind more plausibly contains, namely, the principle of resemblance. (2) Positing this more general principle within our theory of mind allows TCE theory, not only to predict that we will judge that experiences cannot appear nonveridically, but also to predict that we will judge them to be self-intimating. The idea that phenomenal states are self-intimating seems to be so well confirmed by introspection that it features, in some version, in a vast number of philosophical theories: Descartes, Locke, Brentano, Husserl and Chisholm naturally come to mind. More recently, a number of analytic philosophers have defended similar views (BonJour 2000; Fumerton 1995; Kriegel 2009). For this reason, I take it that a correct theory of phenomenal introspection should predict that phenomenal states will come out as self-intimating. Therefore, I think it is an advantage of my new view that it makes such a prediction.

²⁴ Given the way in which the self-intimating character of consciousness and the infallibility of introspection have been derived from the principles of our naïve theory of mind, it should be clear to readers that close variants of the TCE could allow for very similar derivations. For example, one could consider variants of the TCE in which the Principle of Resemblance is replaced by another principle, which does not appeal to the relation of resemblance, but to another relation R, as soon as relation R satisfies some relevant structural constraint (notably: (i) Relation R must relate a state x, which will necessarily be a receptive affection,

Before closing, let us consider some objections. According to TCE theory, we can think of introspective *appearances of experiences*. This is crucial to our proposed solution to the illusion meta-problem, because it is when we think of introspective appearances of experiences that we are led to think that these appearances cannot be nonveridical—which explains our unique resistance to illusionism. But many philosophers have claimed that such second-order appearances are nowhere to be found in our stream of consciousness: there is no such thing as an *introspective phenomenology* (Lycan 1996; Shoemaker 1994; Siewert 2012) and anytime we try to think of an introspective appearance of an experience, we fail and simply end up thinking about the experience itself.

However, I do not think that remarks of this kind are at odds with TCE theory—quite the contrary. Indeed, it is true that TCE theory predicts that we can think of introspective appearances of experiences; but it precisely states that, when we do so, we are unable to isolate these appearances of experiences from the experiences they are the appearances of. Indeed, the nature of phenomenal concepts is such that we necessarily conceive of experiences as entities that are *self-intimating* (there cannot be an experience without it appearing to the subject) and as entities which *cannot appear nonveridically* (there cannot be an appearance of experience without the experience really being there). This has the consequence that any attempt at introspectively focusing on a hypothetical introspective phenomenology (properties of experiences by way of which they appear to the subject) distinct from perceptual phenomenology (properties of experiences by way of which they make sensible states of affairs appear to the subject) is doomed to fail, given that our naïve theory of mind states that the very same properties of experiences, conceived of as receptive affections, do these two jobs at the same time: the same properties of experiences make them *resemble* sensible states of affairs (thus making them appear to the subject) and make them *resemble themselves* (thus making themselves appear to the subject). Therefore, the striking introspective absence of a distinctive *phenomenology of introspection*, far from being a problem for TCE theory, is predicted by it.

Here is another possible objection. TCE theory states that our naïve theory of mind crucially conceives of experiences as resembling the sensory states of affairs they make appear. Most contemporary philosophers deny that such a resemblance obtains; as I noted, this is no problem for the TCE, as the TCE, conjoined with eliminativism, also denies that such resemblance really obtains. However, the fact that most modern and contemporary philosophers (at least since Galileo and Descartes) deny that our experiences resemble external states of affairs (as they state that a red object, for example, does not really bear a quality that resembles our experiences of red) seems

Footnote 24 continued

and a state y , which can be a receptive affection or not (ii) If state y is itself a receptive affection, any state x which enters in relation R with y will be a receptive affection of the same type; (iii) Every given receptive affection x will stand in the relation R to a receptive affection of the same type as itself). Our intuitive relation of resemblance happens to satisfy these constraints, and a concept of resemblance arguably features amongst our basic and intuitive concepts, which is why I formulated TCE using the Principle of Resemblance. However, a philosopher who likes the way in which TCE predicts that experiences will be conceived of as self-intimating, and infallibly introspectively grasped, but does not like its presupposition that our naïve theory of mind contains the Principle of Resemblance, could take inspiration from it and build alternative theories with alternative principles (instead of the Principle of Resemblance) appealing to relations which respect these structural constraints.

to show that such a denial can be easily entertained. But that might seem like a problem for the TCE: after all, if the Principle of Resemblance really is a part of our naïve theory of mind, it should not be so easy to deny that the relevant resemblance holds.²⁵

My answer is this. I think that what many of us easily deny since Galileo and Descartes is not that our experiences present to us the instantiation of simple, primitive qualities (such as colors) by external objects—qualities which really resemble our experiences. What we easily deny is that *there really are such simple, primitive qualities* instantiated in the external world. We take it that *these qualities appear to exist* (because our experiences present them as existing) but that they are not instantiated as such in real external objects. However, the fact that we can do that does not mean we have stopped using the Principle of Resemblance, as we still intuitively think of experiences as resembling the primitive qualities that they present external objects as possessing (though we judge that objects do not really possess such qualities).²⁶ Therefore, I do not think that our easy intuitive acceptance of the absence of *resemblance* between experiences and *actual states of affairs* can be the basis of an argument against the TCE.

Finally, another possible objection is that even if TCE theory delivers a solution to the illusion meta-problem, I have not provided an independent case for the truth of TCE theory. In response, I confess that this is something I have not done here. As stressed above, here the task I have set myself is different: to provide a *how possibly explanation* of our problematic intuitions regarding consciousness, i.e., to show that TCE theory, if true, would deliver a solution to the biggest challenge facing illusionism about consciousness.

What general picture of consciousness and introspection is delivered by the TCE theory? In my view, phenomenal consciousness does not exist—we never are in phenomenal states. However, whenever we focus on our internal states, our theory of mind module automatically applies phenomenal concepts—which is why it *seems to us* that we are phenomenally conscious (and “seeming” here has to be understood in a purely functional sense: we undergo a momentary and non-cognitively penetrable disposition to believe that we are phenomenally conscious). Because our theory of mind is modular (and informationally encapsulated), such seeming (in the functional sense) is resistant to any contradictory beliefs we may acquire: even convinced illusionists are *under the impression* that they are conscious. And, because of the specific nature

²⁵ Thanks to an anonymous reviewer for raising this objection.

²⁶ I am aware of the difficulties that such interpretation raises: my view notably implies that, as soon as we deny that there is a resemblance between experiences and real external objects, we should be intuitively *illusionists* (rather than *reductionists*) about secondary qualities such as colors. But numerous philosophers seem to have easily endorsed reductionist views of such qualities. My answer would be that there is a sense of “quality” (corresponding to the *edenic content of sensory concepts*) such that, as soon as we deny that our experiences really resemble external objects, we will intuitively be illusionists about qualities. However, there may be other motivations to use the word “quality” in a different sense (corresponding to the *ordinary content of sensory concepts*), which would lead one to explicitly defend a *reductionist* view of such qualities. This duality of sense is of course essential from an epistemological point of view: in the first sense of “quality”, post-cartesian philosophers will think that qualities are not instantiated and that our experiences presenting them are all illusory. In the second sense of “quality”, they will think that qualities *are* instantiated, and that our experiences presenting them (in an *indirect* way, mediated by the presentation of primitive qualities which are not really instantiated) are often veridical (which seems crucial to save the epistemological role of experiences).

of the phenomenal concepts then applied (described by the TCE theory), (a) it also automatically seems to us (in the functional sense of seeming) that it seems to us (in the intuitive and innate sense seeming) that we are phenomenally conscious (as we cannot help judging that, if a subject has an experience, it seems to her that she has that experience); (b) we cannot intuitively make sense of our situation regarding phenomenal consciousness as being a situation of illusion; we cannot, without contradiction, think that it *seems to us*—in an intuitive sense—that we are phenomenally conscious while we are not phenomenally conscious (as we cannot help judging that, if it seems to a subject that she has an experience, then she has the experience). Illusionism still is coherent and true, if formulated carefully: it seems to us that we are phenomenally conscious (in the functional sense of seeming), but we are *not* phenomenally conscious, and (crucially) it does *not* seem to us that we are conscious (in the intuitive and innate sense of seeming).²⁷

6 Conclusion

Illusionism as a theory of consciousness presents many theoretical advantages. In particular, it offers a way of defending a physicalist metaphysics without having to solve the hard problem of consciousness. However, it also faces a number of serious problems, the most serious of which is the illusion meta-problem. It stems from the fact that it is uniquely hard to represent to ourselves the state of affairs which illusionism claims is the case. The very idea that experiences appear to exist, but do not really exist, strikes us as puzzling and somewhat absurd, while nothing similar typically happens in the case of other entities. I tried to show that illusionists can account for this fact. I introduced a potential account of our phenomenal concepts, TCE theory, and I argued that, if this theory is true, then we should expect it to be impossible to coherently represent to ourselves the illusory nature of phenomenal consciousness in an intuitive way, even though consciousness really is illusory.

²⁷ An anonymous reviewer pointed out that my view of phenomenal introspection, according to which we conceive of experiences as making states of affairs appear to subjects in virtue of their resemblance with these states of affairs, bears some similarities with U.T. Place's idea according to which we commit a "phenomenological fallacy" about consciousness: Place states that we mistakenly suppose that when a subject "describes his experience, when he describes how things look, sound smell, taste of feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen" (Place 1956, pp. 49–50). This fallacy [that Chalmers Chalmers (2018, pp. 29–30) analyzes as a particular case of what Avenarius (1891) called "introjection"] is, according to Place, what explains our intuitive resistance to materialism regarding consciousness. However, there are important differences between my view and Place's view. First, I think that we are not led to think that there is such resemblance between experiences and the states of affairs they make appear because we commit a fallacy (a kind of reasoning mistake), but because of some inescapable low-level and modular features of our introspective mechanisms [and I argued against attempts at explaining our problematic intuitions regarding consciousness as resulting from a fallacy elsewhere (Kammerer 2018b)]. Second, I think that the idea according to which experiences make states of affairs appear to subjects in virtue of their resemblance with these states of affairs crucially creates problematic intuitions regarding consciousness *because it leads us to ascribe a special epistemological status to consciousness* (self-intimation, introspective infallibility). This point is crucial in my account, while it does not play a similar role in Place's view.

Acknowledgements I want to thank Uriah Kriegel for his extremely valuable comments and remarks. I also want to thank Keith Frankish, Rob Rupert and Samuel Webb for their help and their comments, as well as two anonymous reviewers.

Funding Funding was provided by ANR (Grant Nos. ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL*) and Fonds De La Recherche Scientifique—FNRS.

References

- Armstrong, D. (1980). *The nature of mind and other essays*. Ithaca, NY: Cornell University Press.
- Avenarius, R. (1891). *Der Menschliche Weltbegriff*. Leipzig: Reiland.
- Aydede, M., & Güzeldere, G. (2005). Cognitive architecture, concepts, and introspection: An information-theoretic solution to the problem of phenomenal consciousness. *Noûs*, 39(2), 197–255.
- Balog, K. (2012). Acquaintance and the mind-body problem. In C. Hill & S. Gozzano (Eds.), *New perspectives on type identity: The mental and the physical*. Cambridge: Cambridge University Press.
- Balog, K. (2016). Illusionism's discontent. *Journal of Consciousness Studies*, 23(11–12), 40–51.
- Bayne, T., & Montague, M. (Eds.). (2011). *Cognitive phenomenology*. Oxford: Oxford University Press.
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, 108(1), 1–46.
- BonJour, L. (2000). Toward a defense of empirical foundationalism. In M. DePaul (Ed.), *Resurrecting old-fashioned foundationalism* (pp. 21–38). Lanham, MD: Rowman and Littlefield.
- Carruthers, P. (1996). Simulation and self-knowledge: A defense of theory-theory. In P. Carruthers & P. Smith (Eds.), *Theories of theories of mind* (pp. 22–68). Cambridge: Cambridge University Press.
- Carruthers, P. (2005). *Consciousness: Essays from a higher-order perspective*. Oxford: Clarendon Press.
- Carruthers, P. (2011). *The opacity of mind*. Oxford: Oxford University Press.
- Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers, D. (2002). Consciousness and its place in nature. In D. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings*. Oxford: Oxford University Press.
- Chalmers, D. (2006). Perception and the fall from Eden. In T. Szabo & J. Hawthorne (Eds.), *Perceptual experience* (pp. 49–125). Oxford: Oxford University Press.
- Chalmers, D. (2007). Phenomenal concepts and the explanatory gap. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. Oxford: Oxford University Press.
- Chalmers, D. (2010). *The character of consciousness*. Oxford: Oxford University Press.
- Chalmers, D. (2018). The meta-problem of consciousness. *Journal of Consciousness Studies*, 25(9–10), 6–61.
- Churchland, P. (1979). *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.
- Demircioğlu, E. (2013). Physicalism and phenomenal concepts. *Philosophical Studies*, 165(1), 257–277.
- Dennett, D. (1988). Quining Qualia. In A. Marcel & E. Bisiach (Eds.), *Consciousness in modern science*. Oxford: Oxford University Press.
- Dennett, D. (1991). *Consciousness explained*. London: Penguin.
- Dennett, D. (2017). *From bacteria to bach and back*. New York: Norton & Company.
- Díaz-León, E. (2014). Do a posteriori physicalists get our phenomenal concepts wrong? *Ratio*, 27(1), 1–16.
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge: MIT Press.
- Elpidorou, A. (2013). Having it both ways: Consciousness, unique not otherworldly. *Philosophia*, 41(4), 1181–1203.
- Elpidorou, A. (2016). A posteriori physicalism and introspection. *Pacific Philosophical Quarterly*, 97(4), 474–500.
- Feyerabend, P. (1963). Mental events and the brain. *Journal of Philosophy*, 40, 296–296.
- Frith, C. (2002). Attention to action and awareness of other minds. *Consciousness and Cognition*, 11, 181–187.
- Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language*, 13(1), 1–22.

- Fodor, J. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, Massachusetts: MIT Press.
- Frances, B. (2008). Live skeptical hypotheses. In J. Greco (Ed.), *The Oxford handbook of skepticism* (pp. 225–244). Oxford: Oxford University Press.
- Frankish, K. (2012). Quining diet qualia. *Consciousness and Cognition*, 21(2), 667–676.
- Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, 23(11–12), 11–39.
- Fumerton, R. (1995). *Metaepistemology and skepticism*. Lanham: Rowman and Littlefield.
- Gazzaniga, M. (1970). *The bisected brain*. New York: Appleton Century Crofts.
- Gertler, B. (2008). In defence of mind-body dualism. In J. Feinberg & R. Shafer-Landau (Eds.), *Reason & responsibility: Readings in some basic problems of philosophy* (pp. 285–297). California: Thomson Wadsworth.
- Goff, P. (2011). A posteriori physicalists get our phenomenal concepts wrong. *Australasian Journal of Philosophy*, 89(2), 191–209.
- Goff, P. (2016). Is realism about consciousness compatible with a scientifically respectable worldview? *Journal of Consciousness Studies*, 23(11–12), 83–97.
- Goff, P. (2017). *Consciousness and fundamental reality*. Oxford: Oxford University Press.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.
- Gopnik, A., & Meltzoff, A. (1994). Minds, bodies and persons: Young children's understanding of the self and others as reflected in imitation and « theory of mind » research. In S. Parker, R. Mitchell, & M. Boccia (Eds.), *Self-awareness in animals and humans* (pp. 166–186). New York: Cambridge University Press.
- Graziano, M. (2013). *Consciousness and the social brain*. Oxford: Oxford University Press.
- Graziano, M. (2016). Consciousness engineered. *Journal of Consciousness Studies*, 23(11–12), 98–115.
- Hanson, N. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.
- Hardin, C. L. (1988). *Color for philosophers*. Indianapolis: Hackett.
- Hayes, P. (1978). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the micro-electronic age*. Edinburgh: Edinburgh University Press.
- Hill, C. (1991). *Sensations: A defense of type materialism*. Cambridge: Cambridge University Press.
- Humphrey, N. (2011). *Soul dust: The magic of consciousness*. Princeton: Princeton University Press.
- Humphrey, N. (2016). Redder than red: Illusionism or phenomenal surrealism? *Journal of Consciousness Studies*, 23(11–12), 35–55.
- Humphrey, N. (2017). The invention of consciousness. *Topoi*. <https://doi.org/10.1007/s11245-017-9498-0>.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(April), 127–136.
- Kammerer, F. (2016a). Conscious experiences as ultimate seemings: Renewing the phenomenal concept strategy. *Argumenta*, 1(2), 233–243.
- Kammerer, F. (2016b). *Le problème de l'expérience consciente: une tentative de dissolution*. Ph.D. dissertation, Université Paris-Sorbonne, Paris.
- Kammerer, F. (2016c). The hardest aspect of the illusion problem—And how to solve it. *Journal of Consciousness Studies*, 23(11–12), 123–139.
- Kammerer, F. (2018a). Can you believe it? Illusionism and the illusion meta-problem. *Philosophical Psychology*, 31(1), 44–67.
- Kammerer, F. (2018b). Does the explanatory gap rest on a fallacy? *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-018-0424-1>.
- Kammerer, F. (2018c). Is the antipathetic fallacy responsible for the intuition that consciousness is distinct from the physical? *Croatian Journal of Philosophy*, 18(1), 59–73.
- Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford: Oxford University Press.
- Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(October), 354–361.
- Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford: Oxford University Press.
- Levine, J. (2007). Phenomenal concepts and the materialist constraint. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. Oxford: Oxford University Press.

- Lewis, D. (Ed.). (1983a). Mad pain and Martian pain. In *Philosophical papers* (Vol. I, pp. 122–130). Oxford: Oxford University Press.
- Lewis, D. (Ed.). (1983b). Postscript to “Mad pain and Martian pain”. In *Philosophical papers* (Vol. I). Oxford: Oxford University Press.
- Loar, B. (1997). Phenomenal states (revised version). In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness* (pp. 597–616). Cambridge: MIT Press.
- Lycan, W. (1996). *Consciousness and experience*. Cambridge: MIT Press.
- Maund, B. (2006). The illusion theory of colour: An anti-realist theory. *Dialectica*, 60, 245–268.
- McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. Stevens (Eds.), *Mental models*. Mahwah: Lawrence Erlbaum Associates.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(October), 435–450.
- Nagel, T. (2012). *Mind and cosmos: Why the materialist neo-darwinian conception of nature is almost certainly false*. Oxford: Oxford University Press.
- Nichols, S., & Stich, S. (2003). How to read your own mind: A cognitive theory of self-consciousness. In Q. Smith & A. Jolic (Eds.), *Consciousness: New philosophical perspectives*. Oxford: Oxford University Press.
- Nida-Rümelin, M. (2007). Grasping phenomenal properties. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. Oxford: Oxford University Press.
- Nida-Rümelin, M. (2016). The illusion of illusionism. *Journal of Consciousness Studies*, 23(11–12), 160–171.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
- Pereboom, D. (2011). *Consciousness and the prospects of physicalism*. Oxford: Oxford University Press.
- Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 44–50.
- Reuter, K. (2013). Inferring sensory experiences. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1217–1222). Austin, TX: Cognitive Science Society.
- Reuter, K. (2014). The importance of intentions in introspections. In K. Garvey (Ed.), *J.L. Austin on language* (pp. 197–211). Palgrave Macmillan.
- Rey, G. (1995). Towards a projectivist account of conscious experience. In T. Metzinger (Ed.), *Conscious experience*. Ferdinand Schöningh: Paderborn.
- Rey, G. (2007). Phenomenal content and the richness and determinacy of colour experience. *Journal of Consciousness Studies*, 14(9), 112–131.
- Rorty, R. (1965). Mind-body identity, privacy and categories. *Review of Metaphysics*, 19, 24–54.
- Scholl, B., & Leslie, A. (1999). Modularity, development and « theory of mind ». *Mind and Language*, 14(1), 131–153.
- Schroer, R. (2010). Where’s the beef? Phenomenal concepts as both demonstrative and substantial. *Australasian Journal of Philosophy*, 88(3), 505–522.
- Searle, J. (1997). *The mystery of consciousness*. New York: The New York Review of Books.
- Shea, N. (2014). Using phenomenal concepts to explain away the intuition of contingency. *Philosophical Psychology*, 27(4), 553–570.
- Shoemaker, S. (1994). The phenomenal character of experience. *Philosophy and Phenomenological Research*, 54(2), 291–314.
- Siewert, C. (2012). On the phenomenology of introspection. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness*. Oxford: Oxford University Press.
- Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language*, 20(2), 296–302.
- Strawson, G. (1994). *Mental reality*. Cambridge, Massachusetts: MIT Press.
- Sturgeon, S. (1994). The epistemic view of subjectivity. *The Journal of Philosophy*, 91(5), 221–235.
- Yablo, S. (1999). Concepts and consciousness. *Philosophy and Phenomenological Research*, 59(2), 455–463.